

# NCCS Data Management Plan (DMP)

## Guided Template Version 2.1

- A Description of the project, including:
  - Project Name This should be short, unique, and relevant. Use existing project names where possible. Examples: Landsat (curated data on centralized storage), GISS Modes of Variability (describes owner organization and the science):
  - Principal Investigator This should be the project PI or the person responsible for the data or science:
  - Curator The Curator owns the data online and is responsible for data migration, quality assessment, providing metadata for search and discovery, and future updates:
  - Computational Project ID (if assigned) Relevant for Discover projects with an allocation, the PI or User Services can provide this number:
  - Project Status (On-going/Directed Funding or Finite/Grant Funding) Examples - GMAO, GIS, LIS projects are likely on-going directed funding, ROSES projects are grant funded and finite:
  - Brief Description of work to be completed 1-2 sentences describing the project's purpose:
  - Brief Description of the input, intermediate, and final data files, e.g. observational data, climate model output, in-situ data, etc.
    - Input data are (data generated by another project that is read by your project's code, examples: Landsat, NGA, MODIS):
    - Intermediate files are (data that is created during the computational runs that are not final data products, examples: restart files, output that is post processed):
    - Final Products are (data that are going to be archived, used in publications, are required by a grant):
- Workflow diagrams are optional but strongly encouraged (examples can be provided) (Can show protocols used to transfer data, compute systems that run project code, protocols to move data to an archive)
- Ingest of Input data:
  - Where are the data located now (NASA DAAC, External data service, PI's system, Discover, ADAPT pubrepo, other)?
  - How will the data be brought into the NCCS (network, already available, physical media (special dispensation required))?
  - What is the estimated volume of input data (can be estimated by the size of one file multiplied by the number of files)?
- Access and Sharing:
  - Classification of data:
    - Input Data (Public, Private, Controlled (e.g. ITAR, NGA)):

- Intermediate Files (Public, Private, Controlled (e.g. ITAR, NGA)):
  - Final Products (Public, Private, Controlled (e.g. ITAR, NGA)):
- **Group Access:**
  - Input Data Group (Computational Project, Existing NCCS group (provide group name), Requesting new NCCS group):
  - Intermediate Files (Computational Project, Existing NCCS group (provide group name), Requesting new NCCS group):
  - Final Products (Computational Project, Existing NCCS group (provide group name), Requesting new NCCS group):
  - If a New NCCS group is required, please suggest a name and list the users who will need to be members:
  - Does the data need to be group writable (Y/N)?
  - Does the data need to be world readable (Y/N)?
- **High Level Computational Requirements:**
  - Does your workflow require parallel computation, either traditional HPC with message passing or intrinsic parallelization (Y/N)?
  - NCCS Systems that will need to access the data, if known, specify all that apply (Discover, ADAPT, specific ADAPT VMs, Dataportal):
  - Licensed Software, select all that apply (IDL, Matlab, None, list any other):
  - Open source software requirements, if any:
- **Data Volume (estimated):**
  - Intermediate files that need to be saved during computation and for final QC:
  - Final Products:
  - If the data is to be stored on Centralized Storage, please suggest a directory name:
- **Sharing via Data Services:**
  - Will the data be made available to the public via the NCCS Dataportal (Y/N)?
    - If so, which services (HTTPS, GDS, TDS, ESGF, ArcGIS, FLUID, CREATE-V, other)?
- **Disposition**
  - **When will you delete the following data from disk:**
    - Input Data (after processing, after final QC, after # of years, end of project, etc):
      - If not after processing, provide a justification (e.g. sharing with other NCCS projects, need for future reprocessing):
    - Intermediate Files (after each run, after final QC, after # of years, end of project, etc):
      - If not after processing, provide a justification (e.g. sharing with other NCCS projects, need for future reprocessing):
    - Final Products (after external archiving, after # of years, end of project, etc):
      - If not after processing, provide a justification (e.g. sharing with other NCCS projects, need for future reprocessing):

- **Archive:**
  - Where will the final products be archived (NASA DAAC, PIs system, AWS, other)?
  - How will the final products be migrated to the archive or Cloud (automatically, manually, with NCCS's assistance)?
- **Miscellaneous:**
  - What is the process for managing the data of project members who leave while the project is active, e.g. co-worker's job title, job title, manager's job title?